1. The target variable: 'is_canceled'

Reason: It indicates whether a booking was canceled (1) or not (0). My goal is to build a model that can predict the value of this variable based on the values of other independent variables in the dataset.

2. The predictor:

'lead_time': This could be a significant predictor as longer lead times might be associated with a higher likelihood of cancellations.

'deposit_type': Whether a deposit has been made might influence the likelihood of cancellation, with no deposits possibly correlating to a higher cancellation rate.

'previous_cancellations': Guests who have canceled bookings in the past might be more likely to cancel again.

'customer_type': Different types of customers (transient, contract, group) might have different cancellation behaviors.

'adr': The average daily rate might be a factor influencing cancellations, with higher rates potentially leading to more cancellations.

'total_of_special_requests': A higher number of special requests might indicate a more serious intent to stay, potentially correlating with a lower cancellation rate.

'market_segment' and 'distribution_channel': Different market segments and distribution channels might have different cancellation patterns, offering opportunities for targeted interventions to reduce cancellations.

3. Exploration of the dataset:

3.1 Definition of variables

'is_canceled': Binary attribute indicating if the booking was canceled (1) or not (0).

'lead_time': Number of days between the booking date and the arrival date.

'arrival_date_year': Year of arrival.

'arrival_date_month': Month of arrival.

'arrival_date_week_number': Week number of the arrival date.

'arrival_date_day_of_month': Day of the month of the arrival date.

'stays_in_weekend_nights': Number of weekend nights (Saturday or Sunday) the guest stayed or booked.

'stays_in_week_nights': Number of week nights (Monday to Friday) the guest stayed or booked.

'adults': Number of adults.

'Children': Number of children.

'babies': Number of babies.

'meal': Type of meal booked (e.g., Bed & Breakfast (BB)).

'country': Country of origin of the guest.

'market_segment': Market segment designation (e.g., Direct, Corporate).

'distribution_channel': Booking distribution channel (e.g., Direct, Travel Agents/Operators).

'is_repeated_guest': Binary attribute indicating if the guest was a repeated guest (1) or not (0).

'previous_cancellations': Number of previous bookings that were canceled by the customer.

'previous_bookings_not_canceled': Number of previous bookings that were not canceled.

'reserved_room_type': Code for the type of room reserved.

'assigned_room_type': Code for the type of room assigned.

'booking_changes': Number of changes/amendments made to the booking.

'deposit_type': Indication of a deposit made (e.g., No Deposit).

'agent': ID of the travel agent.

'company': ID of the company responsible for the booking.

'days_in_waiting_list': Number of days the booking was on a waiting list before being confirmed.

'customer_type': Type of booking (e.g., Transient).

'adr': Average Daily Rate.

'required_car_parking_spaces': Number of car parking spaces required.

'total_of_special_requests': Number of special requests made by the customer.

'reservation_status': Last reservation status (e.g., Check-Out).

'reservation_status_date': Date of the last status set.

'name': Name of the guest.

'email': Email of the guest.

'phone-number': Phone number of the guest.

'credit_card': Last four digits of the credit card used.

3.2 Data types: float64(4), int64(16), object(16)

3.3 119390 rows and 36 columns.

3.4 Dataset stats:

	count	mean	std	min	1%	5%	10%	25%	50%	75%	99%	max
is_canceled	119390.0	0.370416	0.482918	0.00	0.0	0.0	0.0	0.00	0.000	1.0	1.00	1.0
lead_time	119390.0	104.011416	106.863097	0.00	0.0	0.0	3.0	18.00	69.000	160.0	444.00	737.0
arrival_date_year	119390.0	2016.156554	0.707476	2015.00	2015.0	2015.0	2015.0	2016.00	2016.000	2017.0	2017.00	2017.0
arrival_date_week_number	119390.0	27.165173	13.605138	1.00	2.0	5.0	8.0	16.00	28.000	38.0	53.00	53.0
arrival_date_day_of_month	119390.0	15.798241	8.780829	1.00	1.0	2.0	4.0	8.00	16.000	23.0	31.00	31.0
stays_in_weekend_nights	119390.0	0.927599	0.998613	0.00	0.0	0.0	0.0	0.00	1.000	2.0	4.00	19.0
stays_in_week_nights	119390.0	2.500302	1.908286	0.00	0.0	0.0	1.0	1.00	2.000	3.0	10.00	50.0
adults	119390.0	1.856403	0.579261	0.00	1.0	1.0	1.0	2.00	2.000	2.0	3.00	55.0
children	119386.0	0.103890	0.398561	0.00	0.0	0.0	0.0	0.00	0.000	0.0	2.00	10.0
babies	119390.0	0.007949	0.097436	0.00	0.0	0.0	0.0	0.00	0.000	0.0	0.00	10.0
is_repeated_guest	119390.0	0.031912	0.175767	0.00	0.0	0.0	0.0	0.00	0.000	0.0	1.00	1.0
previous_cancellations	119390.0	0.087118	0.844336	0.00	0.0	0.0	0.0	0.00	0.000	0.0	1.00	26.0
previous_bookings_not_canceled	119390.0	0.137097	1.497437	0.00	0.0	0.0	0.0	0.00	0.000	0.0	3.00	72.0
booking_changes	119390.0	0.221124	0.652306	0.00	0.0	0.0	0.0	0.00	0.000	0.0	3.00	21.0
agent	103050.0	86.693382	110.774548	1.00	1.0	1.0	6.0	9.00	14.000	229.0	390.00	535.0
company	6797.0	189.266735	131.655015	6.00	16.0	40.0	40.0	62.00	179.000	270.0	513.04	543.0
days_in_waiting_list	119390.0	2.321149	17.594721	0.00	0.0	0.0	0.0	0.00	0.000	0.0	75.00	391.0
adr	119390.0	101.831122	50.535790	-6.38	0.0	38.4	50.0	69.29	94.575	126.0	252.00	5400.0
required_car_parking_spaces	119390.0	0.062518	0.245291	0.00	0.0	0.0	0.0	0.00	0.000	0.0	1.00	8.0
total_of_special_requests	119390.0	0.571363	0.792798	0.00	0.0	0.0	0.0	0.00	0.000	1.0	3.00	5.0